芳香羧酸衍生物驱避剂的非线性定量构效关系

李 颗[1], 李向辉[1], 徐西林[2,3], 袁哲明[2,3,*]

(1. 长沙县农业科学研究所, 长沙 410137; 2. 湖南农业大学, 湖南省作物种质创新与资源利用重点实验室, 长沙 410128; 3. 湖南农业大学, 湖南省植物病虫害生物学及防控重点实验室, 长沙 410128)

摘要:【目的】驱避剂可使害虫不敢接近受用者从而保护受用者免遭其害。建立高精度、可解释性强的非线性定量构效关系(quantitative structure-activity relationship, QSAR)模型对设计合成新的高效昆虫驱避剂有重要意义。【方法】基于37个芳香羧酸类化合物对家蝇 *Musca domestica* 的驱避活性,以量子化学计算软件 PCLIENT 获取每一化合物初始描述符,以二元矩阵重排过滤器、多轮末尾淘汰实施特征非线性筛选,以支持向量回归(support vector regression, SVR)建立非线性 QSAR 模型,以 SVR 非线性解释体系分析各保留描述符对驱避活性的影响。【结果】1 542 个初始描述符的 SVR 模型 $F=1.2$,特征筛选后6个保留描述符的 SVR 模型 $F=184.6$,特征筛选对 QSAR 模型精度有重要影响。6个保留分子描述符的重要性依次为 p4BCD > GATS7v > T(O..O) > JGI8 > SssO > nArCONR2。【结论】保留描述符与芳香羧酸类化合物对家蝇驱避活性的非线性关系明显,获得了高精度、普适性强的非线性 SVR-QSAR 模型。

关键词: 驱避剂; 家蝇; 芳香族衍生物; 驱避活性; 非线性; 定量构效关系; 支持向量回归

中图分类号: Q964 **文献标识码:** A **文章编号:** 0454-6296(2014)09-1018-07

Nonlinear quantitative structure-activity relationship of the aromatic carboxylic acid repellents

LI Ke[1], LI Xiang-Hui[1], XU Xi-Lin[2,3], YUAN Zhe-Ming[2,3,*] (1. Agricultural Sciences Institute of Changsha County, Changsha 410137, China; 2. Hunan Provincial Key Laboratory of Crop Germplasm Innovation and Utilization, Hunan Agricultural University, Changsha 410128, China; 3. Hunan Provincial Key Laboratory for Biology and Control of Plant Diseases and Insect Pests, Hunan Agricultural University, Changsha 410128, China)

Abstract:【Aim】Repellent can protect the users by driving target pests away from them. It is important to establish a nonlinear quantitative structure-activity relationship (QSAR) model with high precision and strong interpretation for designing and synthesizing the new insect repellent with higher bioactivity.【Methods】Based on the repellent activities of 37 aromatic carboxylic acid derivatives against the housefly, *Musca domestica*, the initial descriptors were generated with stoichiometry software PCLIENT, and then the binary matrix shuffling filter (BMSF) and worst descriptor elimination multi-round method (WDEM) were successively used to conduct the nonlinear selection for initial descriptors. With the reserved descriptors, a support vector regression (SVR) model was established for the QSAR analysis of these 37 repellent derivatives. The influence of reserved descriptors on repellent activities was further analyzed with SVR interpretation system.【Results】The F-score of SVR model with original 1 542 descriptors was 1.2. However, it was 184.6 with the retained six descriptors after feature screening, indicating that feature screening has important effects on the precision of QSAR model. The importance of six molecular descriptors was as follows: p4BCD > GATS7v > T(O..O) > JGI8 > SssO > nArCONR2.【Conclusion】The nonlinear relationship between reserved descriptors and the repellent activities of aromatic carboxylic acid derivatives against *M. domestica* was remarkable, and a high-performance SVR-QSAR model for repellent derivatives was constructed.

Key words: Repellent; *Musca domestica*; aromatic carboxylic acid; repellency; nonlinear; quantitative structure-activity relationship (QSAR); support vector regression (SVR)

基金项目: 教育部博士点基金项目(20124320110002)

作者简介: 李颗, 男, 1972 年生, 湖南长沙人, 农艺师, 研究方向为植物保护, E-mail: 113817877@qq.com

* 通讯作者 Corresponding author, E-mail: zhmyuan@sina.com

收稿日期 Received: 2014-06-02; 接受日期 Accepted: 2014-08-06

驱避剂可扰乱害虫自然行为，使害虫不敢接近受用者从而保护受用者免遭其害，具用量小、价格低、不污染环境等优点（郝蕙玲等，2006；Katritzky *et al.*，2008）。在兽医临床上，以已有同类结构驱避剂为基础，设计合成新的高效昆虫驱避剂对预防外寄生虫与昆虫疾病有重要意义（薛飞群等，1997；廖圣良等，2012a）。

定量构效关系（quantitative structure-activity relationship，QSAR）是化学与生物学的桥梁（钟国华和胡美英，2001）。对一组生物活性已知的同类结构化合物，QSAR 建模包括 4 个关键步骤：1）分子描述符获取。可通过量子化学软件计算，对每一化合物获取尽可能全面的分子描述符（Natarajan *et al*，2008）。2）描述符选择。无关与冗余描述符影响建模精度，增加模型复杂性并使得模型解释困难（代志军等，2011）。分子描述符与生物活性间往往存在复杂的非线性关系，常用的逐步线性回归特征选择方法失效（钟国华和胡美英，2001）。本室前期基于支持向量机（support vector machine，SVM）发展了高维特征非线性选择新方法二元矩阵重排过滤器（binary matrix shuffling filter，BMSF）与低维特征非线性选择新方法多轮末尾淘汰（worst descriptor elimination multi-roundly，WDEM），在基于芯片数据的癌信息基因选择、多肽定量序效建模中获得成功应用（代志军等，2011；Zhang *et al.*，2012）。3）回归模型选择。描述符（特征）选择完成后，由于多元线性回归模型非线性解析能力不足，而人工神经网络模型基于经验风险最小，存在不适于小样本、易产生过拟合等弊端，本文选用基于结构风险最小、非线性、适于小样本、能有效避免过拟合的支持向量回归（support vector regression，SVR）建模（Vapnik，1995）。4）模型解释。SVM 缺乏一个显性的表达式，可解释性差。本室前期基于 *F* 测验，对 SVR 建立了一套较完整的非线性解释性体系，并经多个多因素多水平实验设计与配方优化实验验证了其合理性与有效性（李俊等，2010；周世豪等，2012；戴长庚等，2013）。

目前昆虫驱避剂构效关系研究多以吸血蚊虫为靶标，集中于避蚊胺（N，N-diethyl-3-methyl benzamide，DEET）及其类似物（王宗德等，2008）。对 DEET 及其类似物共 40 个化合物的研究表明，影响驱避活性的主要因素包括沸点（蒸汽压或挥发度）、分子的形状大小与亲脂性（疏水性）等（Suryanarayana *et al.*，1991；Katritzky *et al.*，2006）。从驱避机理上，驱避剂或者干扰嗅觉系统以阻断昆虫对宿主气味的识别，或者激活嗅觉神经元引起昆虫的主动躲避行为（廖圣良等，2012a）；因此，驱避剂构效关系需结合昆虫嗅觉感受器相关研究成果。现已证实，DEET 能阻断冈比亚按蚊 *Anopheles gambiae* 嗅觉感受神经对引诱化合物的电生理反应，阻断食物气味对果蝇的引诱行为反应，高度保守的气味受体蛋白 OR83b 是 DEET 的分子靶标（Ditzen *et al.*，2008）。一般认为，从蠕虫到人类，许多气味信号的接受由属于 G 蛋白耦联受体家族中的气味受体与挥发性的配体结合来完成；但新发现昆虫有异源气味受体（Sato *et al.*，2008），昆虫的气味受体形成了配体门控通道和循环核苷激发的无选择性阳离子通道（Wicher *et al.*，2008）。最近，Oliferenko 等（2013）从 43 个酰胺类似物出发，以埃及伊蚊 *Aedes aegypti* 的气味结合蛋白 AaegOPB1 为靶标，通过分子场拓扑分析、分子对接等筛选到了多个有潜力的高驱避活性化合物。

本研究以 37 个芳香羧酸类化合物对家蝇 *Musca domestica* 的驱避活性为对象（薛飞群等，1997），每一化合物以量子化学计算软件 PCLIENT 与文献获取 1 542 个初始描述符（薛飞群等，1999），经 BMSF 与 WDEM 非线性筛选，获得 6 个保留分子描述符，建立了高精度的非线性 SVR-QSAR 模型，进一步以 SVR 非线性解释体系分析了各保留描述符对驱避活性的影响。结果报道如下。

1　材料与方法

1.1　化合物及其驱避活性数据来源

37 个芳香羧酸类化合物及对应的对家蝇的驱避活性指标见表 1（薛飞群等，1999），活性指标为百分驱避率等于 50% 时化合物的摩尔浓度的倒数（$\log 1/C_{50}$）。原文有 40 个化合物，但有 3 个化合物驱避活性未测定，因此舍去。

1.2　分子描述符获取

每一化合物用 PCLIENT 软件（http://www.vcclab.org/lab/pclient/start.html）的 JME 编辑器画出分子结构并导入任务窗口，根据分子结构信息可算得 1 533 个分子描述符；沸点参数（B）、疏水性参数（logP）、电性参数（δ^0）、立体参数（MR_1 和 MR_2）、分子连接性指数（^1X，^2X，^3X 和 $^1X^V$）等 9 个分子描述符引自文献（薛飞群等，1999），其中化合物 M_7 的沸点参数值原文缺失，本文用多元线性回归方法插

表 1 37 个化合物的保留描述符及对家蝇的生物活性

Table 1 The retained descriptors of 37 compounds and their biological activities against *Musca domestica*

化合物 Compounds	保留的描述符 Retained descriptors						生物活性($log1/C_{50}$) Biological activity	
	JGI8	GATS7v	T(O..O)	nArCONR2	p4BCD	SssO	Expe	Pred
B1	0	0	-0.365	0	0	26.30	1.83	1.83
B2	0	0	-0.596	0	0	26.30	2.18	2.13
B3	0	0	-0.685	0	-1.000	26.30	2.53	2.83
B4	0.011	0	-0.930	0	-1.000	26.30	2.55	2.58
B5	0	0	-0.664	0	0	20.31	2.78	2.65
B6	0	0	-0.792	0	-0.923	20.31	3.15	3.06
B7	0.008	0	-0.387	0	-0.007	26.30	1.38	1.72
B8	0	0	-0.785	0	-0.505	20.31	3.80	3.55
E1	0	0	-0.497	0	-0.960	35.53	2.85	2.92
E2	0	0	-0.715	0	-0.883	35.53	3.19	3.11
E3	0	0	-0.790	0	-0.554	35.53	3.22	3.23
E4	0.010	0.385	-1.049	0	-0.262	35.53	3.25	3.39
E5	0	0	-0.789	0	-0.830	29.54	3.19	2.94
E6	0	0	-1.054	0	-0.477	29.54	3.25	3.17
E7	0.007	0.523	-0.486	0	0	35.53	2.68	2.34
E8	0.008	0.448	-0.992	0	-0.035	29.54	3.27	2.95
M1	0	0	-0.240	0	-1.000	35.53	2.52	2.51
M2	0	0	-0.457	0	-0.964	35.53	2.86	2.86
M3	0	0	-0.583	0	-0.790	35.53	3.19	3.16
M4	0.011	0.417	-0.804	0	-0.630	35.53	3.22	3.13
M5	0	0	-0.532	0	-0.907	29.54	2.86	2.88
M6	0	0	-0.846	0	-0.691	29.54	2.91	3.33
M7	0.009	0.540	-0.286	0	0	35.53	2.05	2.33
M8	0.010	0.547	-0.763	0	-0.049	29.54	2.64	3.08
P1	0	0	-0.676	0	0	26.30	2.18	2.30
P2	0.020	0	-0.854	0	-1.000	26.30	2.51	2.54
P3	0.020	0	-0.944	0	-0.991	26.30	2.55	2.51
P4	0.007	0	-1.224	0	-0.733	26.30	2.89	3.04
P5	0	0	-0.592	0	-1.000	20.31	2.81	2.75
P6	0.020	0	-0.888	0	-0.691	20.31	2.88	2.78
P7	0.004	0	-0.645	0	0	26.30	2.03	1.99
P8	0.008	0	-0.759	0	-0.049	20.31	2.60	2.60
S1	0	0	-0.212	1	0	46.53	2.18	1.77
S2	0	0	-0.412	1	-1.000	46.53	2.52	2.52
S3	0	0	-0.468	1	-0.991	46.53	2.56	2.57
S4	0.008	0.460	-0.727	1	-0.649	46.53	3.19	3.18
S7	0.011	0.615	-0.238	1	-0.007	46.53	2.63	2.44

Expe：实验值 Experimental value；Pred：预测值 Predicted value；C_{50}：驱避率等于 50% 时化合物的摩尔浓度 Median repellent molarity.

值补齐。每一化合物共计 1 542 个初始描述符。

1.3 分子描述符选择

二元矩阵重排过滤器 BMSF 高维特征粗筛算法如下：对数据矩阵(y_i, x_{ij}), $i=1,2,\cdots,n$, $j=1,2,\cdots,m$，有 m 个分子描述符，n 个样本。每个分子描述符有 0(不选取)和 1(选取)两种状态。产生一个 $K\times m$ 随机矩阵(本文取 $K=500$)，其元素为 0 或 1，限定每列 0 与 1 的个数相等。从随机矩阵的每行选取值为 1 的矩阵元素，找出原始训练集中对应分子描述符，以 SVR 经 10 折交叉测试获得 K 个均方误差(mean square error, MSE)值。$K\times m$ 随机矩阵(自变量)与 K 个 MSE(因变量)组成新训练集并训练建模，随机矩阵的某列元素 0 和 1 互换后(其他列不变)为测试集，预测得 K 个 MSE_0 与 K 个 MSE_1，若均值 $\overline{MSE_0} \leqslant$ 均值 $\overline{MSE_1}$，则剔除相应分子描述符；遍历 m 次，得第一轮保留分子描述符。重复以上过程，经多轮选择至没有分子描述符可剔除为止(Zhang *et al.*, 2012)。

多轮末尾淘汰 WDEM 低维特征精细筛选算法如下：假定 BMSF 初筛后有 m' 个分子描述符。对数据矩阵(y_i, x_{ij}), $i=1,2,\cdots,n$, $j=1,2,\cdots,m'$，以 SVR 经 10 折交叉测试得初始 MSE_0，第一轮依次去

除第 j 个分子描述符，SVR 交叉测试得对应 MSE_j，若 $\min(MSE_j) \leqslant MSE_0$，则剔除相应分子描述符并进入下一轮筛选，反之筛选结束（代志军等，2011）。假定 WDEM 精筛后有 m'' 个保留分子描述符。

1.4　模型解释

基于 m'' 个保留分子描述符和全部样本构建 SVR 模型。模型非线性回归显著性测验的 F 统计量与决定系数 R^2 由下式给出：

$$F = \frac{U/m''}{Q/(n - m'' - 1)} \tag{1}$$

$$U = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \tag{2}$$

$$Q = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3}$$

$$R^2 = 1 - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{4}$$

其中，U 为回归平方和，Q 为剩余离差平方和，n 为样本数，m'' 为保留分子描述符数，y_i 和 $\hat{y}_i$ 分别为第 i 个样本的真值和估计值，$\bar{y}$ 为所有样本真值的均值，F 的自由度为（m''，$n - m'' - 1$）。若 $F > F_\alpha(m'', n - m'' - 1)$，则在 α 水平上非线性回归显著（谭显胜等，2009）。

单因子重要性显著性测验：固定描述符 x_j 为 $\bar{x}_j$，由预测值可得 U_j 和 Q_j。在 SVR 模型中，$SSy \neq Q + U$，$SSy \neq Q_j + U_j$，其中 $SSy = \sum_{i=1}^{n} (y_i - \bar{y})^2$ 为离差平方和。注意到对同一描述符 x_j，U_j 和 Q_j 的大小仅具相对性，为使各因子间重要性具可比性，可将 Q_j，U_j，Q 和 U 等规格化到 $SSy = Q_j' + U_j' = Q' + U'$，再作 F 测验：

$$Q_j' = Q_j/(Q_j + U_j) \times SSy \tag{5}$$

$$U_j' = U_j/(Q_j + U_j) \times SSy \tag{6}$$

$$Q' = Q/(Q + U) \times SSy \tag{7}$$

$$U' = U/(Q + U) \times SSy \tag{8}$$

$$V_j = U' - U_j' = Q_j' - Q' \tag{9}$$

$$F_j = \frac{V_j/1}{Q'/(n - m'' - 1)} \tag{10}$$

F_j 的自由度为（$1, n - m'' - 1$）（谭显胜等，2009）。

单因子效应分析：将除 x_j 外的各描述符固定为其均值，令 x_j 在给定取值区间内按一定步长取值，代入 SVR 模型得预测值 $\hat{y}_i$，各描述符通过 $x'_j = \dfrac{x_j - \min x}{\max x - \min x}$ 得归一化的 x 轴坐标值，作出 $x_j - \hat{y}_i$ 图（谭显胜等，2009）。

本文 BMSF 高维特征粗筛、WDEM 多轮末尾淘汰精筛、SVR 建模和非线性解释体系等采用自编 MATLAB 程序通过调用 LIBSVM3.1 软件包实现（Chang and Lin，2011）。核函数采用径向基核，核函数参数采用 Python 默认范围、步长经格点搜索自动获取。

2　结果与分析

2.1　芳香羧酸类驱避剂的 SVR-QSAR 模型

基于 37 个样本，1 542 个初始描述符的 SVR 模型 $F = 1.2$，经 BMSF 非线性高维特征初筛后 21 个描述符的 SVR 模型 $F = 4.9$，再经 WDEM 精筛后 6 个保留描述符的 SVR 模型 $F = 184.6 > F_{0.01}(6, 30)$，$R^2 = 0.9731$，非线性回归达极显著。可见特征筛选效果明显。

薛飞群等（1997）从 40 个样本出发，舍弃 5 个含有苯环取代基的化合物和 5 个含有哌啶基的化合物，得到的最优回归方程为：

$$\log 1/C = -1.036 + 0.008B + 0.734\log P - 0.101 \log P^2 + 0.020MR_1 + 0.038MR_2 - 0.176I$$

$n = 30$，$R^2 = 0.9274$，$F = 92.0$。显然，本文所建非线性 SVR 模型精度更高，覆盖样本更多，普适性更强。

模型评估从宽松到严格依次为回代拟合、交叉验证［其极限是留一法（leave one out，LOO）］、独立测试（廖圣良等，2012b）。因样本较少，不划分独立测试集。本文进一步以 LOO 给出了基于 6 个保留分子描述符的 SVR 模型的预测值（表 1，图 1），其 $R^2_{Loo} = 0.8404$。由图 1 可见，预测值与实验值分布于坐标轴对角线附近，进一步表明本文所建非线性 SVR 模型可信。

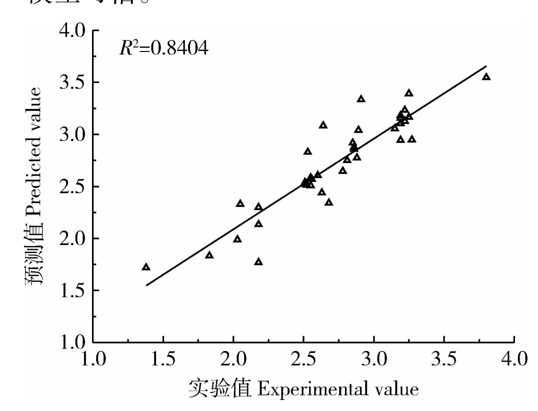


图 1　留一法检验的驱避剂生物活性实验值与预测值

Fig. 1　Experimental values and predicted values of biological activities of repellents with leave-one-out test

2.2 保留分子描述符及其单因子效应

6 个保留分子描述符的单因子重要性显著性测验结果表明，其 F 值均大于临界值 $F_{0.01/6}(1, 30) = 7.56$，达极显著(表 2)。其单因子效应分析结果见图 2，可见 6 个保留分子描述符对驱避活性影响的重要性依次为 p4BCD > GATS7v > T(O..O) > JGI8 > SssO > nArCONR2。保留描述符与驱避活性的非线性关系明显，GSFRAG 程序计算的是分子图像 G 中顶点 $k = 2, 3, \cdots, 10$ 时某一特定片段出现次数，其中 p4BCD 表示片段[片段结构图]，与驱避活性呈开口向下抛物线变化，过高或过低[片段结构图]片段数均对活性不利；GATS7v 是以原子范德华体积加权、步长为 7 时的二维吉尔里自相关值，T(O..O)是两个氧原子间的拓扑距离和，均与驱避活性呈开口向下抛物线变化；SssO 是电性拓扑态的双键氧原子数，与驱避活性呈开口向上抛物线变化；JGI8 为轨道 8 平均拓扑电荷指数，与驱避活性呈近似线性负相关；nArCONR2 为叔酰胺数，化合物中出现叔酰胺对活性不利。在 37 个样本保留描述符取值范围内，当 p4BCD = -0.6122，GATS7v = 0.4267，T(O..O) = -0.9555，JGI8 = 0，SssO = 20.31，nArCONR2 = 0 时化合物预期驱避活性最高。

表 2 特征筛选后的 6 个保留描述符

Table 2 Six retained descriptors after feature screening

序号 No.	组名 Group name	描述符 Descriptor	F 值 F value
1	GSFRAG	p4BCD：GSFRAG 程序算得的分子描述符 p4BCD	571.3**
2	2D autocorrelations	GATS7v：吉尔里自相关值(步长 7，原子范德华体积加权) Geary autocorrelation of lag 7 weighted by van der Waals volume	341.4**
3	2D atom pairs	T(O..O)：两个氧原子拓扑距离和 Sum of topological distances between O..O	272.2**
4	2D autocorrelations	JGI8：轨道 8 平均拓扑电荷指数 Mean topological charge index of order 8	243.7**
5	Atom-type E-state indices	SssO：电性拓扑态的双键氧原子数 Sum of ssO E-states	209.5**
6	Functional group counts	nArCONR2：叔酰胺(芳香胺)数 Number of tertiary amides (aromatic)	172.3**

** $P < 0.01$.

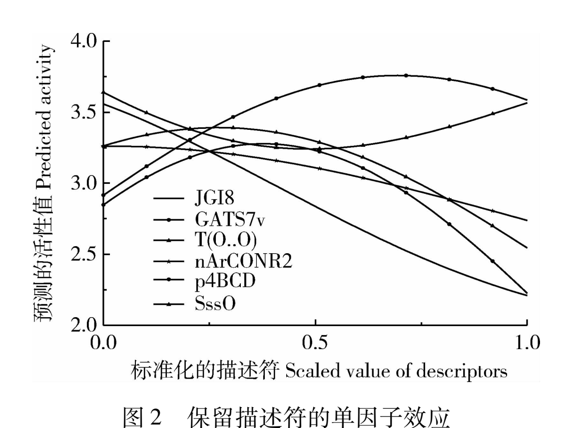


图 2 保留描述符的单因子效应

Fig. 2 Single-factor effects of 6 retained descriptors

3 结论与讨论

分子描述符与生物活性间往往存在复杂的非线性关系。对 DEET 及其类似物共 40 个化合物，Bhonsle 等(2007)给出了一个包含 30 个描述符的较优多元线性回归模型，其中最重要的 6 个描述符依次为相对疏水表面积 Jurs-RASA、表面电场区域参数 Jurs-FPSA-3、Balaban 拓扑指数 JX、ADME 溶解水平(指明分子的液体溶解能力)、分子面积投影指数 Shadow-Xlength 和拓扑描述符 Kappa-3-AM，而摩尔折射率影响不大；尽管其 R^2 高达 0.989，但显然包含保留描述符过多。对芳香羧酸类化合物驱蝇活性，相比薛飞群等(1997)报道的最优六元(拟)线性回归模型($n = 30$)，本文所建立的 6 个保留描述符的非线性 SVR 模型明显精度更高、覆盖样本更多($n = 37$)；图 2 也显示多个保留描述符与驱避活性的单因子效应呈抛物线变化。这表明，QSAR 研究中应优先选用基于结构风险最小、非线性、适于小样本、能有效避免过拟合的 SVR 为基本建模工具。

通常的 2D-QSAR 研究中，化合物分子描述符仅涉及沸点参数(B)、疏水性参数(logP)、电性参数(δ^0)、立体参数(MR_1 和 MR_2)、分子连接性指数(^1X，^2X，^3X 和 $^1X^V$)等少数几种(王宗德等，2008；Garcoa-Domenech *et al.*，2010)，不能全面表征化合物与活性间的复杂关系，常导致建模时需剔除部分“离群”样本(薛飞群等，1997；Wang *et al.*，2008)；本文结果表明，有些“离群”样本，很可能是描述符与建模工具选择不当所致。因此，通过量子化学软件，对每一化合物获取尽可能全面的数以千计初始分子描述符是较为稳健的策略。

然而，小样本、高维特征不但导致“维数灾难”，且无关与冗余描述符影响建模精度，增加模型复杂性并使得模型解释困难，此时特征选择变得尤为关键。本文结果再次证实本室前期发展的高维特征非线性选择新方法 BMSF 与低维特征非线性选择新方法 WDEM 是有效的（代志军等，2011；Zhang *et al.*，2012）。

本文所建立的 37 个样本、6 个保留描述符的 SVR-QSAR 模型为新的蝇类芳香羧酸衍生物高效驱避剂分子设计奠定了基础。未来可进一步搭建多个芳香羧酸衍生物虚拟分子，通过 PCLIENT 量子化学计算在线获取虚拟化合物的 6 个保留描述符，代入模型预测，取预测活性最高且大于 3.8（原数据集 B8 化合物活性最高为 3.8）的少量虚拟化合物，真实合成并实验验证其驱避活性。

参考文献（References）

Bhonsle JB, Bhattacharjee AK, Gupta RK, 2007. Novel semi-automated methodology for developing highly predictive QSAR models: application for development of QSAR models for insect repellent amides. *J. Mol. Model*, 13(1): 179–208.

Chang CC, Lin CJ, 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3): 27.

Dai CG, Li KL, Wang LF, Tan XS, Hu Y, Yuan ZM, Fu Q, 2013. An oligidic diet for *Sesamia inferens* optimized by uniform design and successive rearing. *Chin. J. Rice Sci.*, 27(4): 434–439. [戴长庚，李凯龙，王立峰，谭显胜，胡阳，袁哲明，傅强，2013. 基于均匀设计优化的大螟实用饲料配方及继代饲养. 中国水稻科学，27(4)：434–439]

Dai ZJ, Zhou W, Yuan ZM, 2011. A novel method of nonlinear rapid feature selection for high dimensional data and its application in peptide QSAR modeling based on support vector machine. *Acta Phys. Chim. Sin.*, 27(7): 1654–1660. [代志军，周玮，袁哲明，2011. 基于支持向量机的高维特征非线性快速筛选与肽 QSAR 建模. 物理化学学报，27(7)：1654–1660]

Ditzen M, Pellegrino M, Vosshall LB, 2008. Insect odorant receptors are molecular targets of the insect repellent DEET. *Science*, 319 (5871): 1838–1842.

Garcoa-Domenech R, Aguilera J, Moncef AE, Pocovi S, Gálvez J, 2010. Application of molecular topology to the prediction of mosquito repellents of a group of terpenoid compounds. *Molecular Diversity*, 14 (2): 321–329.

Hao HL, Deng XJ, Du JW, 2006. Extraction of catnip essential oil components and their repellent activity against *Aedes albopictus* and *Culex pipiens pallens*. *Acta Entomologica Sinica*, 49 (3): 533–537. [郝蕙玲，邓晓军，杜家纬，2006. 猫薄荷精油有效成分的提取及其对白纹伊蚊、淡色库蚊的驱避活性. 昆虫学报，49 (3)：533–537]

Katritzky AR, Dobchev DA, Tulp I, Karelsonc M, Carlson DA, 2006. QSAR study of mosquito repellents using Codessa Pro. *Bioorg. Med. Chem. Lett.*, 16(8): 2306–2311.

Katritzky AR, Wang ZQ, Slavov S, Tsikolia M, Dobchev D, Akhmedov NG, Hall CD, Bernier UR, Clark CC, Linthicum KJ, 2008. Synthesis and bioassay of improved mosquito repellents predicted from chemical structure. *Proc. Natl. Acad. Sci. USA*, 105(21): 7359–7364.

Li J, Tan XS, Tan SQ, Yuan ZM, Xiong XY, 2010. Application of improved support vector machine in the optimization of artificial diet for the cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae). *Acta Entomologica Sinica*, 53(4): 420–426. [李俊，谭显胜，谭泗桥，袁哲明，熊兴耀，2010. 改进支持向量机在棉铃虫人工饲料配方优化中的应用. 昆虫学报，53(4)：420–426]

Liao SL, Jiang ZK, Song J, Wang ZD, Han ZJ, Chen JZ, 2012a. Repelling mechanism of mosquitoes repellent. *Chinese Journal of Hygienic Insecticides & Equipments*, 18(4): 280–283. [廖圣良，姜志宽，宋杰，王宗德，韩招久，陈金珠，2012a. 蚊虫驱避剂的驱避机理研究. 中华卫生杀虫药械，18(4)：280–283]

Liao SL, Song J, Wang ZD, Chen JZ, Chen SX, Fan GR, Jiang ZK, Han ZJ, 2012b. Quantitative calculation of the influence of the molecular association between terpenoid repellents and CO_2 on their repellency against mosquitoes. *Acta Entomologica Sinica*, 55(9): 1054–1061. [廖圣良，宋杰，王宗德，陈金珠，陈尚钘，范国荣，姜志宽，韩招久，2012b. 定量计算萜类驱避化合物与二氧化碳缔合对其蚊虫驱避活性的影响. 昆虫学报，55(9)：1054–1061]

Natarajan R, Basak SC, Mills D, Kraker JJ, Hawkins DM, 2008. Quantitative structure-activity relationship modeling of mosquito repellents using calculated descriptors. *Croatica Chemica Acta*, 81 (2): 333–340.

Oliferenko PV, Oliferenko AA, Poda GI, Osolodkin DI, Pillai GG, Bernier UR, Tsikolia M, Agramonte NM, Clark GG, Linthicum KJ, Katritzky AR, 2013. Promising *Aedes aegypti* repellent chemotypes identified through integrated QSAR, virtual screening, synthesis, and bioassay. *PLoS ONE*, 8(9): e64547.

Sato K, Pellegrino M, Nakagawa T, Nakagawa T, Vosshall LB, Touhara K, 2008. Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature*, 452(7190): 1002–1006.

Suryanarayana MVS, Pandey KS, Prakash S, Raghuveeran CD, Dangi RS, Swamy RV, Rao KM, 1991. Structure-activity relationship studies with mosquito repellent amides. *Journal of Pharmaceutical Sciences*, 80(11): 1055–1057.

Tan XS, Wang ZM, Tan SQ, Yuan ZM, Xiong XY, 2009. Establishing interpretability for support vector regression. *Journal of System Simulation*, 21(24): 7795–7797. [谭显胜，王志明，谭泗桥，袁哲明，熊兴耀，2009. 支持向量回归可解释性体系的建立. 系统仿真学报，21(24)：7795–7797]

Vapnik VN, 1995. The Nature of Statistical Learning Theory. Springer Verlag Press, New York. 87–189.

Wang ZD, Song J, Chen JZ, Song ZQ, Shang SB, Jiang ZK, Han ZJ,

2008. QSAR study of mosquito repellents from terpenoid with a six-member-ring. *Bioorganic & Medicinal Chemistry Letters*, 18(9): 2854-2859.

Wang ZD, Song J, Jiang ZK, Han ZJ, Chen JZ, Song ZQ, Shang SB, Chen C, 2008. Study of the structure-activity relationship and mechanism of repellent. *Chinese Journal of Hygienic Insecticides & Equipments*, 14(6): 472-476. [王宗德, 宋杰, 姜志宽, 韩招久, 陈金珠, 宋湛谦, 商士斌, 陈超, 2008. 驱避剂的构效关系和驱避机理的研究. 中华卫生杀虫药械, 14(6): 472-476]

Wicher D, Schafer R, Bauernfeind R, Stensmyr MC, Heller R, Heinemann SH, Hansson BS, 2008. *Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature*, 452(7190): 1007-1011.

Xue FQ, Wang HQ, Zhao RC, 1997. Determination of repellency of aromatic carboxylic acid derivatives to housefly in relationship to quantitative structure activity. *Scientia Agricultura Sinica*, 30(1): 77-83. [薛飞群, 汪汉卿, 赵荣材, 1997. 芳香羧酸衍生物驱避剂活性测定及定量构效关系的研究. 中国农业科学, 30(1): 77-83]

Xue FQ, Wang YC, Xu ZZ, 1999. Quantitative structure activity relationship study with the aromatic acid repellents by molecular connectivity. *Chinese Journal of Veterinary Drug*, 33(2): 12-15. [薛飞群, 王玉春, 徐忠赞, 1999. 分子连接性法分析芳香羧酸类驱避剂的构效关系. 中国兽药杂志, 33(2): 12-15]

Zhang HY, Wang HY, Dai ZJ, Chen MS, Yuan ZM, 2012. Improving accuracy for cancer classification with a new algorithm for genes selection. *BMC Bioinformatics*, 13(1): 298.

Zhong GH, Hu MY, 2001. QSAR and its application and advance in pesticide design. *Chinese Journal of Pesticide Science*, 3(2): 1-11. [钟国华, 胡美英, 2001. QSAR 及其在农药设计中的应用和进展. 农药学学报, 3(2): 1-11]

Zhou SH, Li J, Yao RX, Zhang X, Yuan ZM, 2012. Optimization of chemically defined diet for larvae of the cotton bollworm (*Helicoverpa armigera*) based on uniform design and support vector regression. *Acta Entomologica Sinica*, 55(1): 124-132. [周世豪, 李俊, 姚润贤, 张星, 袁哲明, 2012. 基于均匀设计与支持向量回归的棉铃虫幼虫全纯人工饲料配方优化. 昆虫学报, 55(1): 124-132]

(责任编辑: 赵利辉)